80-29

Optimal Robust Designs for Some Regression Problems

by

W. Notz[1]
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-29

[1] Research supported by National Science Foundation Grant NSF-MCS-7901707.

Optimal Robust Designs For Some
Regression Problems

by

W. Notz[1]
Purdue University

Abstract. This paper deals with finding optimal designs for linear models where the regression function is of the form $\vec{\theta}'\vec{f}(\vec{x})$ + bias. $\vec{f}(\vec{x})$ is assumed to be known but little is known about the bias other than that it is bounded in absolute value for all $\vec{x}$ by some constant c and is such that the linear model is well defined. It is assumed that $\vec{\theta}$ will be estimated by the usual least squares estimates. Optimal designs for estimating $\vec{\theta}$ when $\vec{f}(\vec{x}) = (1, \vec{x}')'$ are found for a number of optimality criteria. When $\vec{f}(x) = (1, x, x^2, \ldots, x^n)'$ the analog of D-optimal designs for estimating $\vec{\theta}$ are found. In all cases considered it is shown that the "usual" optimal designs are optimal for the model with bias considered here. This indicates that the usual optimal designs are "robust" when the regression function is not known precisely.

Key words: Optimum designs, Φ_p-optimality, D-optimality, robustness.

[1]Research supported by National Science Foundation Grant NSF-MCS-7901707.

1. Introduction

In this paper we shall be considering experimental designs for regression models of the form

$$(1.1) \qquad Ey(\vec{x}) = \beta_0 + \vec{\beta}'\vec{x} + \psi(\vec{x})$$

where $\vec{x} \in X$ = some subset of k-dimensional Euclidean space, β_0 is a real number, $\vec{\beta}$ is some vector in k-dimensional Euclidean space, and $\psi \in G$. G is a set of real-valued functions on X having the properties

(i) $|\psi(\vec{x})| \leq c$ for a fixed constant c, for all $\vec{x} \in X$ and all $\psi \in G$.

(ii) For any finite set of points $\vec{x}_1,\ldots,\vec{x}_n$ in X and real numbers $a_1,\ldots,a_n$ such that $|a_i| \leq c$ for all i, there is a $\psi \in G$ such that $\psi(\vec{x}_i) = a_i$ for all i.

(iii) β_0 and $\vec{\beta}$ are well-defined in (1.1).

An example of such a G is given in section 2.

For any finite set of points $\vec{x}_1,\ldots,\vec{x}_n$ we assume the observed values of y at these points, namely $y(\vec{x}_1),\ldots,y(\vec{x}_n)$ are uncorrelated random variables all having variance σ^2. β_0 and $\vec{\beta}$ are unknown constants which will be estimated by the "usual"least squares estimates, i.e. the least squares estimates one would get assuming $\psi = 0$. Our objective is to find an experimental design which will yield the most "accurate" (in some sense) estimates of β_0 and $\vec{\beta}$ assuming model (1.1) holds and we have no knowledge of what "bias" $\psi \in G$ is present. More precisely, we seek designs which will minimize the maximum over all $\psi \in G$ of some function of

$$(1.2) \qquad E(\hat{\vec{\theta}} - \vec{\theta})(\hat{\vec{\theta}} - \vec{\theta})'$$

where $\vec{\theta} = (\beta_0, \vec{\beta}')'$ and $\hat{\vec{\theta}}$ is the least squares estimate of $\vec{\theta}$. In Sections 2, 3, and 4 of this paper minimizing designs are found in various settings. In all cases considered we find that the best designs are the designs one would use if ψ was identically 0 in (1.1). In Section 5 some implications of the results of this paper are discussed and brief reviews of a few other papers dealing with models like (1.1) are given.

2. Φ_p - optimality of the usual designs for linear regression in a robust setting.

Let R^k denote k dimensional Euclidean space and let Γ be the group of all coordinate permutations and reflections in R^k. Let X be the k-fold Cartesian product of the closed interval [-1, +1]. Notice X is invariant under Γ.

Suppose we observe n values $y_1,\ldots,y_n$ of a real-valued random variable at points $\vec{x}(1),\ldots,\vec{x}(n)$ in X respectively. The value of y_i is believed to depend on $\vec{x}(i)$ through the probability model

$$(2.1) \qquad y_i = \beta_0 + \vec{\beta}'\vec{x}(i) + \psi(\vec{x}(i)) + e_i, \quad i = 1,\ldots,n.$$

The e_i are uncorrelated random variables with mean 0 and variance σ^2. Also $\beta_0 \in R$ and $\vec{\beta} \in R^k$ are unknown constants we wish to estimate. We further assume that $\psi \in G$ where G is any set of real valued functions on X having the properties

(i) if $\psi \in G$ then $|\psi(\vec{x})| \leq c$ for some constant $c > 0$ for $\vec{x} \in X$.

(ii) For any finite collection of points $\vec{x}_1,\dots,\vec{x}_m \in X$ and any set of real numbers $a_1,\dots,a_m$ such that $|a_i| \leq c$ for all i, there is a $\psi \in G$ such that $\psi(\vec{x}_i) = a_i$ for each i.

(iii) β_0 and $\vec{\beta}$ are well defined in (2.1).

An example of such a G is the set of all real valued functions ψ on X satisfying $|\psi(\vec{x})| \leq c$ for all $\vec{x} \in X$ and $\int\psi(\vec{x})d\vec{x} = \int x_1\psi(\vec{x})d\vec{x} = \dots = \int x_k\psi(\vec{x})d\vec{x} = 0$ where x_i is the i-th coordinate of $\vec{x}$, $d\vec{x}$ is lebesgue measure on X, and all integrals are over X. Notice for this G the term $\beta_0 + \vec{\beta}'\vec{x}(i)$ in (2.1) represents the best linear approximation to y (considered as a function of $\vec{x}$) in the L_2-norm with respect to lebesgue measure.

Our objective, roughly speaking, will be to decide how best to select $\vec{x}(1),\dots,\vec{x}(n)$ so as to get the most accurate (in some sense) least squares estimates of β_0 and $\vec{\beta}$ in the absence of any information about ψ. This objective will be made more precise as we proceed.

To make our objective more precise let us write the model (2.1) as

(2.2) $$Ey(\vec{x}) = \beta_0 + \vec{\beta}'\vec{x} + \psi(\vec{x})$$

exhibiting the dependence of y on $\vec{x}$. In this paper we shall take an experimental design to be any probability measure on X. If ξ is an experimental design, the information matrix of ξ, denoted $M(\xi)$, is defined in this setting to be

$$M(\xi) = \int_X (1, \vec{x}')'(1, \vec{x}')d\xi(\vec{x})$$

Notice $M(\xi)$ is a nonnegative definite $(k+1)\times(k+1)$ matrix. From here on we shall restrict attention to designs ξ having finite support and such that $M(\xi)$ is nonsingular.

For any $\psi \in G$ define

(2.3) $\psi_0 = \int \psi(\vec{x})d\xi(x), \ \psi_i = \int x_i\psi(\vec{x})d\xi(\vec{x}), \quad i = 1,\ldots,k$

All integrals above and from here on are to be assumed to be over X unless otherwise specified.

Also define

(2.4) $$\vec{\psi} = (\psi_0, \psi_1,\ldots,\psi_k)', \ \vec{\theta} = (\beta_0, \vec{\beta}')'.$$

For any integer $n > 0$ a probability measure ξ having support on points $\vec{x}(1),\ldots,\vec{x}(m) \in X$ ($m \leq n$) and putting mass $r(i)/n$ on $\vec{x}(i)$ ($r(i)$ is an integer $\leq n$ for $i = 1,\ldots,m$) corresponds to a design where one takes a total of n observations, $r(i)$ of them at the point $\vec{x}(i)$ for $i = 1,\ldots,m$. Such a design is called an exact design and is the only type of design which one can implement in practice.

Suppose ξ is an exact design taking n observations, where $n > 0$ is an integer, and $M(\xi)$ is nonsingular. If $\hat{\vec{\theta}}$ is the least squares estimate of $\vec{\theta}$ and we let $\rho = \sigma^2/n$ one can show

$$E(\hat{\vec{\theta}} - \vec{\theta})(\hat{\vec{\theta}} - \vec{\theta})' = \rho M^{-1}(\xi) + M^{-1}(\xi)\vec{\psi}\vec{\psi}'M^{-1}(\xi).$$

We define

(2.5) $$D(\xi, \psi) = \rho M^{-1}(\xi) + M^{-1}(\xi)\vec{\psi}\vec{\psi}'M^{-1}(\xi).$$

The first term on the right hand side of (2.5) is the covariance matrix of $\hat{\vec{\theta}}$. The second term on the right hand side of (2.5) represents the bias present in our estimate $\hat{\vec{\theta}}$ of $\vec{\theta}$ due to the unknown function ψ in our model (2.2). Our objective can now be stated as trying to find a design ξ which minimizes (in some sense) the maximum over all possible ψ of $E(\hat{\vec{\theta}}-\vec{\theta})(\hat{\vec{\theta}}-\vec{\theta})' = D(\xi,\psi)$. In seeking this minimizing design we shall not just restrict attention to exact designs taking n observations but rather allow ξ to be any design having n or fewer points in its support.

To see in what sense ξ will minimize the maximum over ψ of $D(\xi,\psi)$, we define for any $(k+1)\times(k+1)$ positive definite matrix D and any $0 < p < \infty$,

$$\Phi_p(D) = ((\sum_{i=0}^{k} \lambda_i^p)/(k+1))^{1/p} \tag{2.6}$$

where the λ_i are the eigenvalues of D. We seek a ξ which minimizes $\sup \Phi_p(D(\xi,\psi))$ for a given $p \geq 1$, where the sup is over all $\psi \in G$. Notice when $p = 1$ this minimizing ξ corresponds to what is sometimes called an A-optimal design.

To find this minimizing ξ, define for any design ξ

$$m_{ij}(\xi) = \int x_i x_j d\xi(\vec{x}) \qquad 0 \leq i,\ j \leq k \tag{2.7}$$

where $x_0 = 1$. Notice $m_{ij}(\xi)$ is the (i+1) - st, (j+1) - st entry of $M(\xi)$.

We also define $\Xi(m)$ to be the set of all designs ξ having finite support (but not necessarily n or fewer points in the support) and satisfying $m_{00}(\xi) = 1$, $m_{0i}(\xi) = m_{i0}(\xi) = 0$ for all $1 \leq i \leq k$, $m_{ij}(\xi) = 0$ for all $1 \leq i \neq j \leq k$, and $m_{ii}(\xi) = m$ for all $i = 1,\ldots,k$.

Finally we define

$$\lambda(\xi,\psi) = \vec{\psi}' M^{-1}(\xi) M^{-1}(\xi)\vec{\psi}. \tag{2.8}$$

Notice $\lambda(\xi,\psi)$ is the only non-zero eigenvalue of the rank 1 matrix $M^{-1}(\xi)\vec{\psi}\vec{\psi}' M^{-1}(\xi)$ and the eigenvector having eigenvalue $\lambda(\xi,\psi)$ is $M^{-1}(\xi)\vec{\psi}$.

The following lemmas will enable us to find the minimizing ξ.

<u>LEMMA 2.1</u>. If $\xi \in \Xi(m)$ then $c^2 M^{-1}(\xi) - M^{-1}(\xi)\vec{\psi}\vec{\psi}' M^{-1}(\xi)$ is non-negative definte and

$$\lambda(\xi,\psi) \leq (\int \psi^2(x)d\xi)/m \leq c^2/m \tag{2.9}$$

pf. Since $\xi \in \Xi(m)$ an orthonormal set of functions in $L_2(\xi)$ (the L_2-space with respect to the measure ξ on X) is $\{1, x_1/\sqrt{m},\ldots,x_k/\sqrt{m}\}$. The projection of any $\psi \in G$ onto the subspace of $L_2(\xi)$ spanned by this set is

$$(2.10) \qquad \text{Proj}\ \psi(\vec{x}) = \int \psi(\vec{x})d\vec{x} + \sum_{i=1}^{k} \left(\int \psi(\vec{x})(x_i/\sqrt{m})d\vec{x}\right)x_i/\sqrt{m}$$

$$= \psi_0 + \sum_{i=1}^{k} \psi_i x_i/\sqrt{m}$$

Since $\xi \in \Xi(m)$ we have by (2.8) that $\lambda(\xi,\psi) = \psi_0^2 + \sum_{i=1}^{k} \psi_i^2/m^2$. Using this and the fact that Proj ψ has length no greater than ψ in the norm induced by $L_2(\xi)$ (i.e. the length of a function ψ is $\int\psi^2(\vec{x})d\vec{x}$) we have

$$(2.11) \qquad c^2 \geq \int \psi^2(\vec{x})d\vec{x}$$

$$\geq \int (\text{Proj}\ \psi(\vec{x}))^2 d\vec{x}$$

$$= \psi_0^2 + \sum_{i=1}^{k} \psi_i^2/m_i$$

$$= \text{tr}\ \vec{\psi}'M^{-1}(\xi)\vec{\psi}$$

$$= m\lambda(\xi,\psi) + (1 - m)\psi_0^2$$

$$\geq m\lambda(\xi,\psi)$$

since $m \leq 1$. This yields (2.9).

Next since $\text{tr}\ \vec{\psi}'M^{-1}(\xi)\vec{\psi} = \text{tr}\ \vec{\psi}\vec{\psi}'M^{-1}(\xi) \leq c^2$ it follows that the only non-zero eigenvalue of $\vec{\psi}\vec{\psi}'M^{-1}(\xi)$ is $\leq c^2$. From this it follows that $c^2M^{-1}(\xi) - M^{-1}(\xi)\vec{\psi}\vec{\psi}'M^{-1}(\xi)$ is non-negative definite.

LEMMA 2.2. Suppose $\xi \in \Xi(1)$. Let $\pi \in G$ be such that $\pi = c$ on supp ξ. Then for $1 \leq p < \infty$

(2.12) $$\sup_{\psi \in G} \Phi_p(D(\xi, \psi)) = \Phi_p(D(\xi, \pi))$$

pf. For any $\psi \in G$ and $\xi \in \Xi$ (1),

$$D(\xi,\psi) = \rho\, \text{diag}(1, 1, \ldots, 1) + M^{-1}(\xi)\vec{\psi}\vec{\psi}' M^{-1}(\xi)$$

where diag $(a_1, a_2, \ldots, a_{k+1})$ is a $(k+1)\times(k+1)$ diagonal matrix having a_i as the i-th entry on the diagonal. We can write the eigenvalues of this $D(\xi,\psi)$ in the form $\rho + \alpha_0$, $\rho + \alpha_1$, $\rho + \alpha_2$, ..., $\rho + \alpha_k$ for some $\alpha_0, \alpha_1, \ldots, \alpha_k$. By Lemma 2.1 we know $c^2 M^{-1}(\xi) - M^{-1}(\xi)\vec{\psi}\vec{\psi}' M^{-1}(\xi)$ is non-negative definite hence the eigenvalues of $D(\xi,\psi)$ must satisfy $\rho \le \rho + \alpha_0 \le \rho + c^2$, $\rho \le \rho + \alpha_1 \le \rho + c^2$, ..., $\rho \le \rho + \alpha_k \le \rho + c^2$. This means the α_i must satisfy

(2.13) $$0 \le \alpha_0 \le c^2, \quad 0 \le \alpha_1 \le c^2, \ldots, 0 \le \alpha_k \le c^2.$$

Lemma 2.1 also tell us that $\text{tr}\, M^{-1}(\xi)\vec{\psi}\vec{\psi}' M^{-1}(\xi) = \lambda(\xi,\psi) \le c^2$ hence we must have

$$\begin{aligned}\sum_{i=0}^{k} (\rho + \alpha_i) &= \text{tr}\, D(\xi,\psi) \\ &= \text{tr}\, \rho\text{diag}(1, 1, \ldots, 1) + \text{tr}\, M^{-1}(\xi)\vec{\psi}\vec{\psi}' M^{-1}(\xi) \\ &\le \rho + k\rho + c^2\end{aligned}$$

so that the α_i must further satisfy

(2.14) $$\sum_{i=0}^{k} \alpha_i \le c^2.$$

Now

(2.15) $$\Phi_p(D(\xi,\psi)) = [[(\rho + \alpha_0)^p + \sum_{i=1}^{k} (\rho + \alpha_i)^p]/(k + 1)]^{1/p}.$$

Using convexity or Lagrange multipliers one can show for $p \ge 1$ that values of $\alpha_0, \alpha_1, \ldots, \alpha_k$ which maximize $(\rho + \alpha_0)^p + \sum_{i=1}^{k} [\rho + \alpha_i]^p$ subject to

(2.13) and (2.14) are $\alpha_0 = c^2$, $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$. Thus any $\psi^* \in G$ which has the property that the eigenvalues of $D(\xi,\psi)$ are $\rho + c^2, \ldots,$ will satisfy for $1 \leq p < \infty$, $\xi \in \Xi$ (1)

(2.16) $$\sup_{\psi \in G} \Phi_p(D(\xi, \psi)) = \Phi_p(D(\xi, \overset{*}{\psi})).$$

It is easy to check that the π given in the lemma has this property.

<u>LEMMA 2.3</u>. Suppose $\xi^* \in \Xi$ (1), then for any design ξ having finite support and for which $M(\xi)$ is nonsingular

(2.17) $$\sup_{\psi \in G} \Phi_p(D(\xi, \psi)) \leq \sup_{\psi \in G} \Phi_p(D(\xi^*, \psi)), \; 1 \leq p < \infty.$$

pf. For any design ξ having finite support let $\pi(\xi)$ be a fixed element of G having the property that $\pi(\xi)(\vec{x}) = c$ if $\vec{x} \in \sup \xi$. Also for any element γ of the finite group Γ of coordinate reflections and permutations in R^k let $\gamma\xi$ be the design having the property

(2.18) $$\gamma\xi(\vec{x}) = \xi(\gamma\vec{x}), \quad \vec{x} \in X$$

Now for $p \geq 1$ tr M^p is convex in M. Also M^{-1} is convex in nonsingular M. Notice $M^{-1}(\xi)\vec{\pi}(\xi) = (c,0,\ldots,0)'$ since $\vec{\pi}(\xi)$ is the first row of $cM(\xi)$. Since $\sup_{\psi \in G} D(\xi,\psi) = \sup_{\psi \in G} D(\gamma\xi,\psi)$ for all $\gamma \in \Gamma$ we have

(2.19) $$\begin{aligned}\sup_{\psi \in G} \operatorname{tr}[D(\xi, \psi)]^p &= (1/\operatorname{card}\Gamma) \sum_{\gamma\in\Gamma} \sup_{\psi \in G} \operatorname{tr}[D(\gamma\xi, \psi)]^p \\ &\geq (1/\operatorname{card}\Gamma) \sum_{\gamma\in\Gamma} \operatorname{tr}[D(\gamma\xi,\pi(\gamma\xi)]^p \\ &= (1/\operatorname{card}\Gamma) \sum_{\gamma\in\Gamma} \operatorname{tr}[\rho M^{-1}(\gamma\xi) + \operatorname{diag}(c^2,0,\ldots,0)]^p \\ &\geq \operatorname{tr}[(1/\operatorname{card}\Gamma) \sum_{\gamma\in\Gamma} \rho M^{-1}(\gamma\xi) + \operatorname{diag}(c^2,0,\ldots,0)]^p \\ &\geq \operatorname{tr}[\rho M^{-1}((1/\operatorname{card}\Gamma) \sum_{\gamma\in\Gamma} \gamma\xi) + \operatorname{diag}(c^2,0,\ldots,0)]^p\end{aligned}$$

where convexity is used in the last two inequalities.

Let $\eta = (1/\text{card}\Gamma) \sum_{\gamma \in \Gamma} \gamma\xi$. Then η is a design having finite support, $M(\eta)$ is non-singular, and $\eta \in \Xi(b)$ for some $b \leq 1$. Thus

$$(2.20) \qquad \sup_{\psi \in G} \text{tr}[D(\xi, \psi)]^p \geq \text{tr}[\rho M^{-1}(\eta) + \text{diag}(c^2, 0, \ldots, 0)]^p$$

$$\geq \text{tr}[\ M^{-1}(\xi^*) + \text{diag}(c^2, 0, \ldots, 0)]^p$$

$$= \text{tr}[D(\xi^*, \pi(\xi^*))]^p$$

$$= \sup_{\psi \in G} \text{tr}\ D^p(\xi^*, \psi)$$

where the second inequality comes from the fact $b \leq 1$ and the last equality follows from Lemma 2.2 and the fact that $x^{1/p}$ is increasing in $x > 0$ for fixed $p \geq 1$. The lemma now follows from (2.20).

THEOREM 2.1. Consider the set $\Xi(1)$.

Suppose n is an integer > 0 such that there exists an exact design $\xi^* \in \Xi(1)$ taking n observations (thus $D(\xi^*, \psi) = E(\hat{\vec{\theta}} - \vec{\theta})(\hat{\vec{\theta}} - \vec{\theta})'$). Then for all $p \geq 1$

$$\inf_{\xi} \sup_{\psi \in G} \Phi_p(D(\xi, \psi)) = \sup_{\psi \in G} \Phi_p(D(\xi^*, \psi))$$

where the inf is over all designs ξ for which $M(\xi)$ is nonsingular.

pf: This follows from Lemma 2.2.

It is well known that for the model

$$(2.21) \qquad y(x_i) = \beta_0 + \vec{\beta}'\vec{x}_i + e_i$$

where everything is as in equation (2.1), a design ξ on X which will minimize $\Phi_p(M^{-1}(\xi))$ is any $\xi \in \Xi(m)$ where m is as large as possible. In practice, therefore, Theorem 2.1 says that one should go ahead and use the design which is optimum for model (2.21) (a model without a bias term) even if you think your model is really (2.1). Loosely speaking this means that if you have a model like (2.1) but have no idea as to the form of the bias ψ present the best design to use is the one you would choose if no bias ψ whatsoever was present.

It should be remarked that a similar result can be obtained for the setting where X is the simplex in R^k and model (2.1) holds except that $\beta_0 = 0$.

3. D-optimality of the usual designs for linear regression in a robust setting

In this section all notation is the same as in Section 2. Our model is once again (2.1) except now we let X be any subset of the k-fold Cartesian product of the closed interval [-1, +1]. In this section our objective is to find a design ξ which minimizes

$$\sup_{\psi \in G} \det D(\xi, \psi).$$

To carry out this minimization let us introduce some notation. Suppose ξ is a design having support on the n points $\vec{x}(1),\ldots,\vec{x}(n) \in X$. Let $S(\xi)$ be the $(k+1) \times n$ matrix whose i-th column is the vector $(1, \vec{x}(i)')'$. Let $W(\xi)$ be the $n \times n$ diagonal matrix $\mathrm{diag}(\xi(\vec{x}(1)), \xi(\vec{x}(2)), \ldots, \xi(\vec{x}(n)))$. Notice we can write

$$(3.1) \qquad M(\xi) = S(\xi)W(\xi)S'(\xi)$$

$$\vec{\psi} = S(\xi)W(\xi)(\psi(\vec{x}(1)),\ldots,\psi(\vec{x}(n)))'.$$

We get the following result

THEOREM 3.1. Suppose there exists a design ξ^* having k+1 points in its support and such that

(i) $M(\xi^*)$ is nonsingular.

(ii) $\det M^{-1}(\xi^*) = \inf \det M^{-1}(\xi)$, where the inf is over all designs ξ on X.

Then ξ^* minimizes sup det $D(\xi,\psi)$ among all designs ξ on X

pf. If ξ^* is as in the theorem, $S(\xi^*)$ and $W(\xi^*)$ are $(k+1)\times(k+1)$ matrices and it is easy to verify $\xi^*(x) = 1/(k+1)$, if $\vec{x}$ is in the support of ξ^*, in order for ξ^* to minimize $\det M^{-1}(\xi)$ or equivalently maximize $\det M(\xi)$. One than has for any $\psi \in G$

$$(3.2)\quad \det D(\xi^*,\psi) = \det[\rho\, M^{-1}(\xi^*) + M^{-1}(\xi^*)\vec{\psi}\vec{\psi}'M^{-1}(\xi^*)]$$

$$= \det M^{-2}(\xi^*)\, \det[\rho\, M(\xi^*) + \vec{\psi}\vec{\psi}']$$

$$= \det M^{-2}(\xi^*) \cdot \det M(\xi^*) \cdot \rho^k(\rho + \vec{\psi}'M^{-1}(\xi^*)\vec{\psi})$$

$$= \det M^{-1}(\xi^*) \cdot \rho^k(\rho + \vec{\psi}'M^{-1}(\xi^*)\vec{\psi})$$

$$= \det M^{-1}(\xi^*) \cdot \rho^k(\rho + (\psi(\vec{x}(1)),\ldots,\psi(\vec{x}(k+1)))W(\xi^*)S'(\xi^*) \cdot$$

$$S^{-1}{}'(\xi^*)W^{-1}(\xi^*)S^{-1}(\xi^*)S(\xi^*)W(\xi^*)(\psi(\vec{x}(1)),\ldots,\psi(\vec{x}(k+1)))')$$

$$= \det M^{-1}(\xi^*) \cdot \rho^k(\rho + \sum_{i=1}^{k+1} \psi^2(\vec{x}(i))\xi^*(\vec{x}(i)))$$

$$\leq \det M^{-1}(\xi^*)\ \ \rho^k(\rho + c^2).$$

The last inequality will be equality if and only if $\psi = c$ on supp ξ^*. If we let ψ_0 be any element of G such that $\psi_0 = c$ on supp ξ^* then

$$(3.3)\quad \inf_{\xi}\sup_{\psi\in G} \det D(\xi,\psi) \geq \inf_{\xi} \det D(\xi,\psi_0)$$

$$= \inf_{\xi} \det M^{-1}(\xi) \cdot \rho^k(\rho + c^2)$$

$$= \det M^{-1}(\xi^*)\ \ \rho^k(\rho + c^2)$$

$$= \sup_{\psi \in G} \det D(\xi^*, \psi)$$

$$\geq \inf_{\xi} \sup_{\psi \in G} \det D(\xi, \psi).$$

The second line in (3.3) is proved in a manner analogous to (3.2). The third line follows from the choice of ξ^*. The fourth line follows from (3.2). (3.3) yields

$$\sup_{\psi \in G} \det D(\xi^*, \psi) = \inf_{\xi} \sup_{\psi \in G} \det D(\xi, \psi)$$

where the inf is over all designs ξ on X. The theorem is proved.

Since $M(\xi)$ is the same for any design ξ which minimizes $\det M^{-1}(\xi)$ (such a design is called D-optimal), it follows that if there is a D-optimal design on k+1 points of X for the model (2.21) (the usual linear model without bias) any D-optimal design for this model will minimize $\sup_{\psi \in G} \det D(\xi, \psi)$ among all designs ξ on X for model (2.1). Once again we see that if essentially nothing is assumed about the bias in (2.1) the optimal design is the same as for the model where no bias is present, i.e. $\psi = 0$.

4. D-optimality of the usual designs for polynomial regression in a robust setting.

Consider the probability model

$$y_i = \beta_0 + \sum_{r=1}^{k} \beta_r x_i^r + \psi(x_i) + e_i, \quad i = 1, \ldots, n \tag{4.1}$$

where $x \in [-1, +1]$, $\beta_0, \beta_1, \ldots, \beta_k \in R$, the e_i are uncorrelated random variables with mean 0 and variance σ^2, and $\psi \in G$ where G is any set of real valued functions on [-1, +1] having the properties

(i) if $\psi \in G$ then $|\psi(x)| \leq c$ for some constant $c > 0$ and all $x \in [-1, +1]$.

(ii) For any finite collection of points $x_1,\ldots,x_m$ in $[-1, +1]$ and any set of real numbers $a_1,\ldots,a_m$ such that $|a_i| \leq c$ for all i, there is a $\psi \in G$ such that $\psi(x_i) = a_i$ for each i.

(iii) β_0, $\beta_1,\ldots,\beta_{k-1}$, and β_k are all well defined in (4.1).

An example of such a G is the set of all real valued functions ψ on $[-1, +1]$ satisfying $|\psi(x)| \leq c$ for all $x \in [-1, +1]$ and $\int \psi(x)dx = \int x\psi(x)dx = \ldots = \int x^k\psi(x)dx = 0$ where dx is lebesque measure on $[-1, +1]$ and all intergrals are over $[-1, +1]$. Notice for this G the term $\beta_0 + \beta_1 x + \ldots + \beta_k x^k$ is the best approximation to y by a polynomial of degree k in the L_2 norm with respect to lebesque measure.

This model is just a special case of that considered in Section 3 with $\vec{x} = (x, x^2,\ldots,x^k)'$ and

$$X = \{(x, x^2,\ldots,x^k)' \in R^k;\ -1 \leq x \leq +1\}.$$

If we apply the results of Section 3 in this special case we get the following.

<u>THEOREM 4.1</u>. Suppose there exists a design ξ^* having k+1 points in its support and such that

(i) $M(\xi^*)$ is nonsingular.

(ii) $\det M^{-1}(\xi^*) = \inf \det M^{-1}(\xi)$, where the inf is over all designs ξ on $[-1, +1]$.

Then ξ^* minimizes $\sup \det D(\xi,\psi)$ among all designs ξ on $[-1, +1]$.

<u>pf</u>. This is an application of Theorem 3.1.

As a result of this theorem one finds that the D-optimal designs of Hoel (1958) and Guest (1958) for the polynomial regression model

$$y_i = \beta_0 + \sum_{r=1}^{k} \beta_r x_i^r + e_i, \quad i = 1,\ldots,n \tag{4.2}$$

where the notation here is the same as in equation (4.1), are also the designs which minimize $\sup_{\psi \in G} D(\xi,\psi)$ for the model (4.2). This is because the designs of Hoel and Guest have support on exactly k+1 points of [-1, +1] and hence satisfy the conditions of Theorem 4.1. Once again we see that if essentially nothing is known about the bias term ψ in (4.2) one might as well behave as though no bias was present in the model.

5. Discussion

The results of this paper, on the surface, are probably not very interesting or satisfying from a practical standpoint. In the case of simple linear regression on an interval it does not seem reasonable to take only observations on the endpoints of the interval if one is not sure that the underlying model is a straight line. Ones intuition says one ought to include some observations from the interior of the interval. The results of this paper, though, show that there is no good way to choose these interior points for observation if little is known about the departure from linearity of the regression function. One needs to know something about the departure from linearity, which is called the bias in this paper, in order to choose interior points for observation. For example one needs to know the bias is convex, concave, or quadratic. In the absence of such information the "best" one can do is to behave as though no bias is present. What to do when something is known about the form of the bias has been discussed in some special cases by other authors.

There are several papers in the literature dealing with models like (1.1) and it may be instructive to comment on a few of them in closing.

Box and Draper (1959) appear to have been the first to point out the dangers of a large bias term in estimation in any strict formulation of a regression model that ignores the possibility that the chosen model may only be an approximation to the true model. A careful description of some problems in this context is given by Kiefer (1973). Huber (1975) considered a simple linear regression model like (1.1) with $X = [-\frac{1}{2}, +\frac{1}{2}]$ and ψ to be square integrable with respect to lebesgue measure on X. Since ψ is unbounded on any finite set of points in X the bias term can be arbitrarily large and hence no design on a finite set of points has finite risk for the risk function considered by Huber. The optimal design is found to have continuous support and hence not implementable in practice. Marcus and Sacks (1976) formulated a simple linear regression model like (1.1) with $X = [-1, +1]$, bounded by a known function $\phi(x)$, $\psi(0) = 0$, and only linear estimates (not necessarily least squares estimates) of the parameters are considered. This prior knowledge about ψ, i.e. that $|\psi(x)| \leq \phi(x)$ and $\psi(0) = 0$, yields interesting results. Optimal designs depend to some extent on the form of $\phi(x)$. In addition the fact that $\psi(0) = 0$ means no bias is present at $x = 0$ and hence it turns out for many ϕ that $x = 0$ is in the support of the optimal design. Pesotchinsky (1978) has considered multivariate formulations of the Marcus and Sacks model. Notz (1979) and Li and Notz (1980) have considered a model like (1.1) except the estimates of $\vec{\theta}$ are only required to be linear estimates but not necessarily the least squares estimates. Only the mean square error between the estimates and the true values of the unknown parameters is considered. Wu (1977) considers a robust formulation in the spirit of (1.1) for some classical experimental design situations. He comes to conclusions similar to those of Sections 2, 3, and 4 in this paper.

In summary, the results of this paper seem to indicate that in considering models of the form (1.1) with unknown bias one needs to have prior knowledge of or assumptions about the form of the bias otherwise there is no way of knowing where to take observations other than at points which are optimum for the model without bias.

REFERENCES

1. Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54, 622-654.

2. Guest, P. G. (1958). The spacing of observations in polynomial regression. Ann. Math. Statist. 29, 294-299.

3. Hoel, P. G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Statist. 29, 1134-1145.

4. Huber, P. (1975). Robustness and designs. A Survey of Statistical Design and Linear Models, North Holland, Amsterdam, 287-303.

5. Kiefer, J. (1973). Optimal Designs for Fitting Biased Multiresponse Surfaces. Multivariate Analysis III, Academic Press, New York, 287-297.

6. Li, K. C. and Notz, W. I. (1980). Robust designs for nearly linear regression. Dept. of Statistics, Division of Mathematical Sciences, Purdue University Mimeograph Series #80-3.

7. Marcus, M. B. and Sacks, J. (1976). Robust Designs for Regression Problems. Statistical Decision Theory and Related Topics II, Academic Press, New York, 245-268.

8. Notz, W. I. (1979). Robust designs for nearly linear regression. Dept. of Statistics, Division of Mathematical Sciences, Purdue University Mimeograph Series #79-11.

9. Pesotchinsky, L. (1978). Optimal robust designs for linear regression. To appear.

10. Wu, C. F. (1977). Robustness and efficiency problems of some randomization procedures in experimental designs. Univeristy of Wisconsin MRC Technical Summary Report #1797.